

Summary Report: The Effectiveness of Online Ads: A Field Experiment

Alexander Coppock and David Broockman*

September 16, 2015

This document is a summary of experimental findings only. Additionally, this pilot experiment should not be taken as indicative of the effectiveness of online ads more generally.

Abstract

Using a novel field experimental design, we rigorously tested the effectiveness of an online banner ad in changing political attitudes. We find that attitudes towards the #blacklivesmatter campaign's issues were unaffected by a week's worth of exposure ads. This null result cannot be attributed to small sample sizes or noisy outcome measures: Our experiment was powerful enough to detect even small changes in attitudes.

One of the most commonplace forms of political persuasion is the online banner ad. Banner ads typically consist of an image, some short text, and a link to website where those who are interested can get more information. Because clickthrough rates are extraordinarily low, we hypothesize that the mechanism by which banner ads may change attitudes is *not* the information that users gain after clicking through; instead, banner ads are supposed to work by making politicians or issues more salient in the minds of voters.

The study tests this proposition using a randomized field experiment in which some subjects are exposed to ads while others are not. We find that, at least in this experiment, banner ads did not change minds towards a recent political issue. Further, the ads did not even make the issue more salient in subjects' minds.

1 Design

Our experiment proceeded as follows. First, we recruited 1824 subjects were recruited via Amazon's Mechanical Turk to a pre-treatment survey in which demographic covariates were collected.

*The design and analysis of this study was preregistered at egap.org

The cover for this pre-survey was an unrelated survey experiment. Using the built-in random assignment tools in our online survey platform (Qualtrics), we assigned 919 subjects to the treatment group and 909 to control. Over the course of the 10 days between the pre-survey and the post-survey, treatment group subjects were exposed to the advertisement shown in Figure 1. All in all, we served 25,135 ad impressions, generating a total of 18 clicks, for a clickthrough rate of 0.07%. Supposing that all treatment group subjects saw approximately the same number of ads, we estimate that treated subjects were exposed to the advertisement 27 times over the course of the treatment period. We were able to collect outcome data for 1,429 subjects, for a recontact rate of 78.3%. Under the assumption that reporting status is unaffected by treatment, our estimates of the effects of treatment will apply to Always-reporters (those who respond to the post-treatment survey regardless of assignment) only.

Figure 1: Treatment Ad



Anticipating that at best, the banner ad would have small effects, we created an index from 10 post-treatment questions all having to do with the #blacklivesmatter campaign. The questions are presented below and were combined into an index using principal components analysis. This index is our main dependent variable, but we will use the last two questions (Problem and Priority) as dependent variables as well.

1. Recently, statistics have shown that some city police departments are disproportionately likely to use excessive force when arresting black citizens. Some people say this issue is a major problem, although others point out that it's only one of many challenges police departments face. Do you agree or disagree with the statements below? Do you agree or disagree with the statements below? (7 point scale, Strongly disagree - Strongly Agree)
 - (a) Police officers should undergo mandatory training on how to avoid prejudice against

black citizens

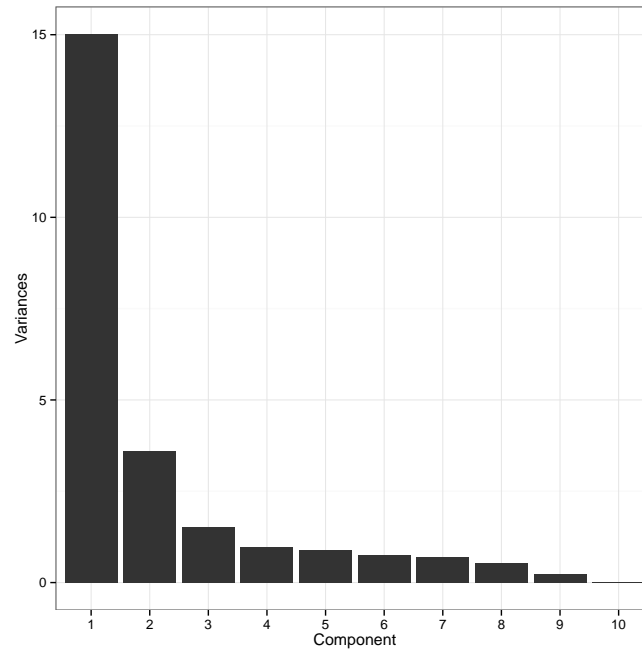
- (b) Politicians should spend more time thinking about how to address racial disparities in who police arrest
 - (c) Statistics that show black drivers are more likely to be stopped suggest that police officers need to be more careful to not act in a biased manner
 - (d) Although the police have discriminated against many groups in the past, today blacks still receive more discrimination than most other groups.
2. Do you agree or disagree with the statements below? (7 point scale, Strongly disagree - Strongly Agree)
- (a) The recent killings of unarmed African American men by police in Ferguson, Missouri and New York City are a sign of broader problems in treatment of African Americans by police
 - (b) Our country's criminal justice system treats blacks and whites equally
 - (c) None of the police officers in my community are prejudiced against blacks
 - (d) Blacks are stopped more frequently by police because they commit more crimes.

DV: Problem How much do you think the mistreatment of Blacks by police is a problem? (4 point scale, Not at all a problem - Serious Problem)

DV: Priority Considering the many other challenges faced by police departments, how high of a priority should addressing the mistreatment of Blacks by police be? (5 point scale, Not a priority to Essential)

Figure 2 shows the scree plot for our principal components analysis of the 10 dependent variables. The plot shows that a majority of the variance can be accounted for using the the first dimension alone; we use the first dimension as our main dependent variable.

Figure 2: PCA Skree Plot



2 Results

Figure 3 shows the distribution of our outcome scale by treatment condition. Both distributions are left-skewed, reflecting the general level of support for #blacklivesmatter among the relatively liberal Mechanical Turk population.

Figure 3: Distribution of Outcome, by condition

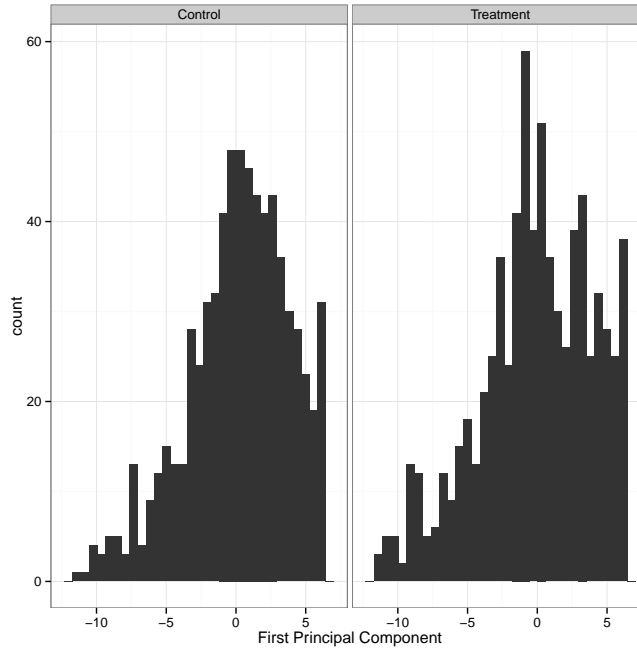


Table 1 shows our main results. We estimate treatment effects via OLS, with and without controls for demographic variables. The covariate-adjusted estimates are slightly more precise, so we will focus on them here. The estimated effect on the scale is -0.113 , with a standard error of 0.179 , a relatively tightly estimated null effect. To put this estimate in perspective, the standard deviation of the scale is 4: the estimated effect is $0.113/4 = 0.028$ standard deviations. Interpretations vary with regard to what constitutes a “small” effect – Cohen (1977) considers an effect size of 0.2 to be small. Our experiment had the power to detect an average movement of approximately half a “small” effect: $2.8 * 0.179/4 = 0.125$ standard deviations.

The other two dependent variables considered here, Problem and Priority, likewise show only very small movements as a result of the treatment.

Table 1: Effects of Online Advertisements on Attitudes towards #Blacklivesmatter

	Scale		Priority		Problem	
Assigned to Advertisements	-0.362	-0.113	-0.068	-0.025	-0.099	-0.040
	(0.204)	(0.179)	(0.054)	(0.051)	(0.048)	(0.043)
Constant	0.187	1.731	3.544	3.591	3.013	3.217
	(0.139)	(0.560)	(0.037)	(0.167)	(0.034)	(0.136)
Covariates	No	Yes	No	Yes	No	Yes
N	1,429	1,429	1,429	1,429	1,429	1,429
R ²	0.002	0.243	0.001	0.125	0.003	0.214

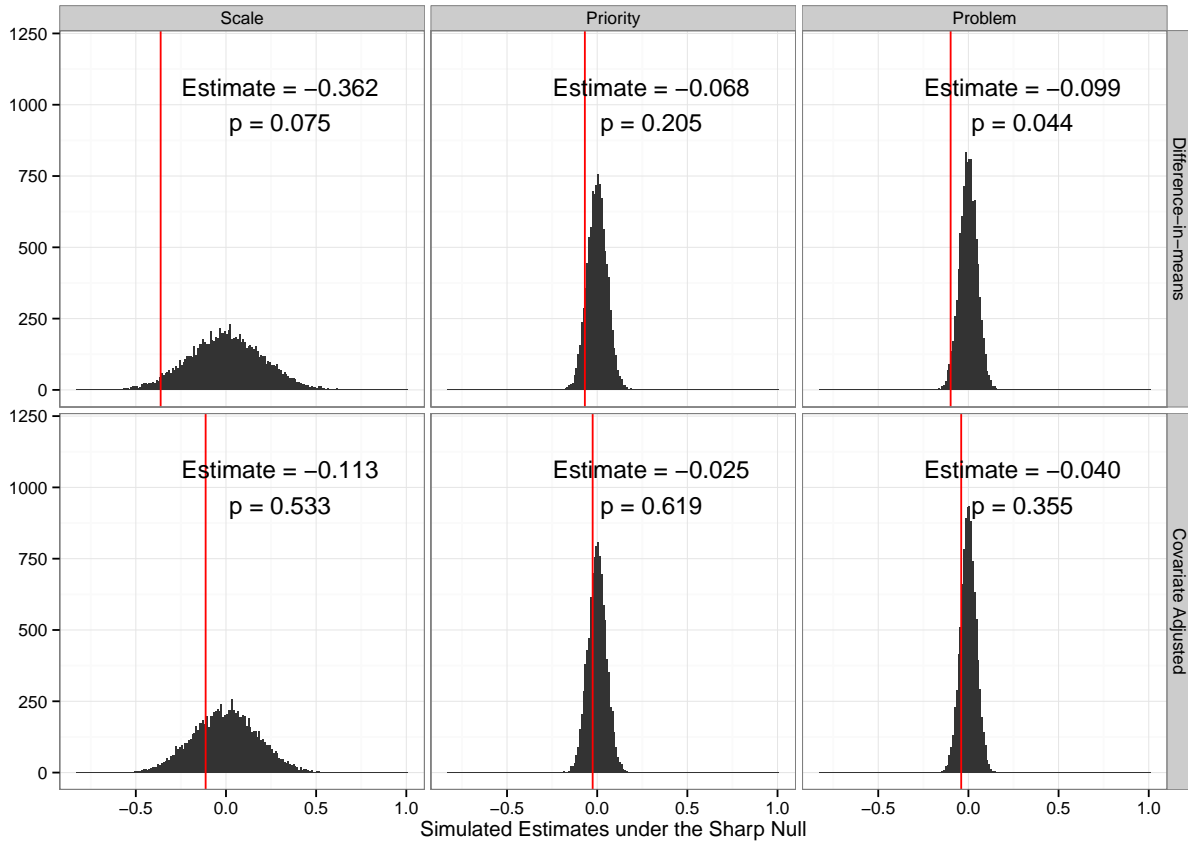
NA

Robust standard errors in parentheses.

Covariates include age, gender, ideology, party id, and race.

We use randomization inference to conduct hypothesis tests. Figure 4 shows the distributions of the estimated effects under the sharp null of no effect for any unit. The top row shows the difference-in-means estimates all our three dependent variables and the bottom row our covariate-adjusted estimates. The two-sided p-values shown in each panel indicate that our estimated differences were all insignificant: We fail to reject the sharp null in all six cases.

Figure 4: Randomization Inference: Simulations under the Sharp Null



3 Checks

In this section, we conduct a number of checks. Table 2 shows the results of a manipulation check to see if the advertising campaign increased familiarity with #blacklivesmatter. Subjects were asked how familiar they were with each of these hashtag campaigns (5 point scale: not at all familiar to very familiar) #getcovered, #JeSuisCharlie, #blacklivesmatter, #IceBucketChallenge, #Race-Together, #BringBackOurGirls, #ilridewithyou, #occupywallstreet, #yesallwomen, #whyistayed, #gamergate. As indicated by the very small coefficients, the treatment did not perceptibly increase familiarity with #blacklivesmatter. We are confident that the treatment group subjects were in fact exposed to the ads – we added ourselves to the same marketing list that controls who sees the adds and did in fact encounter the ads over the course of the treatment period.

Table 2: Manipulation Check

	Manipulation Check	
Assigned to Advertisements	-0.009 (0.027)	-0.002 (0.026)
Constant	0.507 (0.019)	0.657 (0.078)
Covariates	No	Yes
N	1,403	1,403
R ²	0.0001	0.086

NA

Robust standard errors in parentheses.

Covariates include age, gender, ideology, party id, and race.

Our recontact rate was 79.3%: in Table 3 we test to see if reporting status is related to treatment. Treatment group subjects were four percentage points more likely to respond to the post-treatment survey than control group subjects; this difference is significant at the 5% level. We are puzzled as to what the mechanism accounting for this difference could possibly be, but if this estimate reflects the causal impact of the ads on responding to the follow up, then our assumption that we are estimating effects among Always-reporters only is incorrect.

Table 3: Attrition Check

	Responded to Survey	
Assigned to Advertisements	0.040 (0.019)	0.040 (0.019)
Constant	0.763 (0.014)	0.564 (0.060)
Covariates	No	Yes
N	1,824	1,824
R ²	0.002	0.019

NA

Robust standard errors in parentheses.

Covariates include age, gender, ideology, party id, and race.

In Table 4, we present demographic profile of the treatment and control groups and show that they do not differ by more than would be expected by chance. The omnibus p -value presented in the last row comes from a randomization inference procedure in which the F-statistic from an OLS regression of treatment on the covariates is compared to the null distribution of f-statistics over 10,000 possible random assignments. The large p -value indicates that the covariates are not predictive of the treatment indicator, consistent with random assignment.

Table 4: Balance Table

	Treatment Group		Control Group		Difference	
	Mean	SD	Mean	SD	Estimate	SE
Age	4.01	1.15	4.00	1.10	0.01	0.06
Female	0.40	0.49	0.44	0.50	-0.03	0.03
Liberal	0.44	0.25	0.52	0.25	-0.08	0.01
Moderate	0.31	0.21	0.26	0.19	0.04	0.01
Conservative	0.21	0.16	0.18	0.15	0.03	0.01
Ideology: Other	0.04	0.04	0.04	0.04	0.00	0.00
7-point Party ID	3.29	1.87	3.14	1.82	0.15	0.10
Education	3.56	0.85	3.58	0.88	-0.02	0.04
White	0.71	0.21	0.67	0.22	0.03	0.01
Black	0.05	0.05	0.06	0.06	-0.01	0.00
Hispanic	0.06	0.05	0.04	0.04	0.01	0.00
Race: Other	0.19	0.15	0.22	0.17	-0.03	0.01

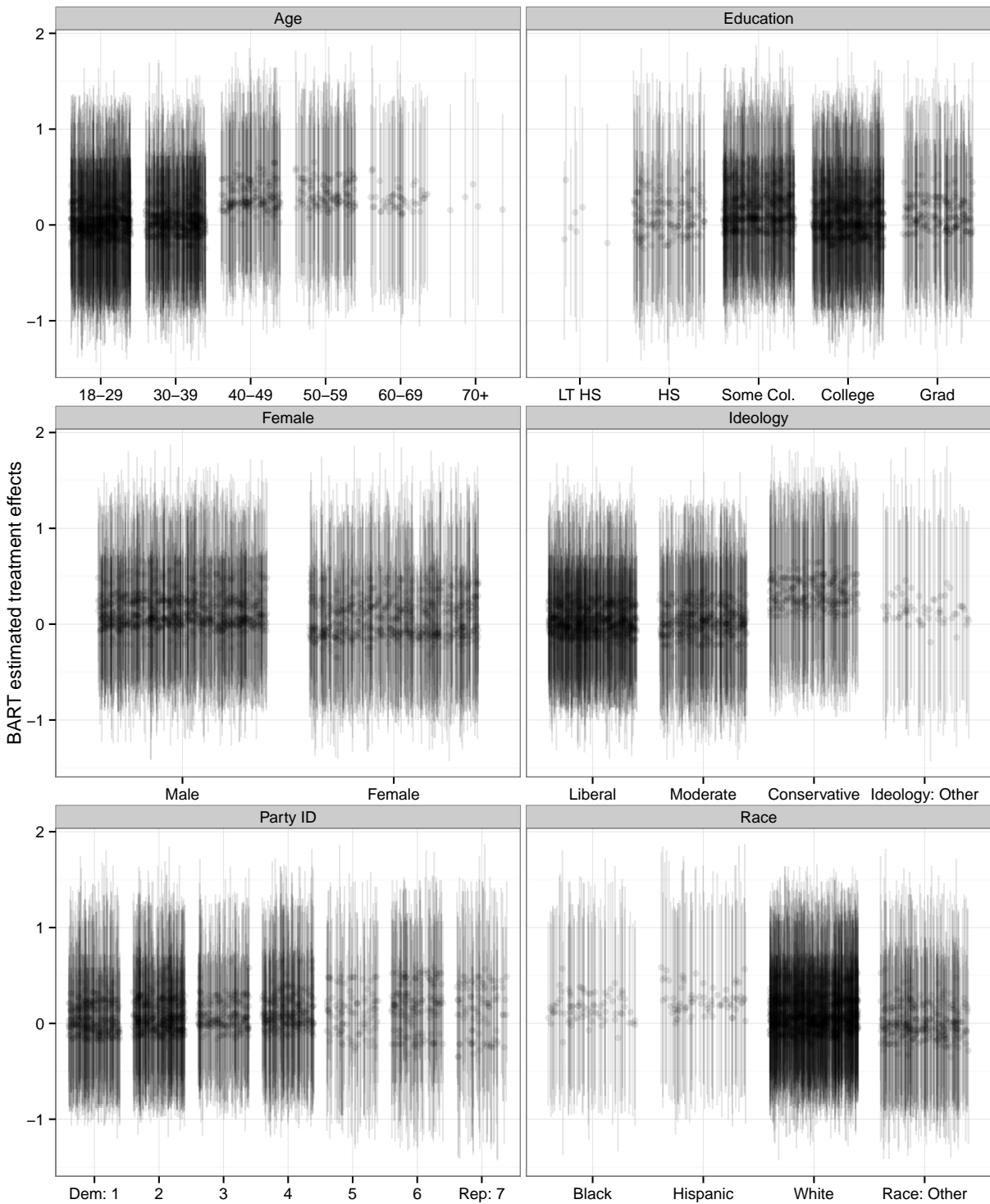
Omnibus F: 1.47
RI p -value: 0.147

4 Exploring Heterogeneity with BART

In our preanalysis plan, we indicated that we would explore treatment effect heterogeneity with Bayesian Additive Regression Trees (BART).

As shown in Figure 5, there is no evidence that the treatment was effective for an particular subgroup, nor is there much evidence of variability in the estimated treatment effects.

Figure 5: BART Estimated Treatment Effects, by Covariates



5 Discussion

This experiment has shows that our display ad was altogether ineffective. It did not change minds nor did it increase the salience of #blacklivesmatter in the minds of our subjects. We cannot attribute this null finding to weak experimental power or to a “bad draw.”

Our null finding raises some questions. Is the ad itself ineffective, or did our subjects simply never pay it any attention? A survey experiment in which subjects are directly asked to view the ad might shed some light on this question. Would a different ad be more effective? Perhaps a candidate-centered advertisement might raise name recognition. We plan to explore these and other questions in future experiments.